

The Stability of Boolean Networks with Generalized Canalizing Rules

Andrew Pomerance,* Michelle Girvan, and Ed Ott

Institute for Research in Electronics and Applied Physics

University of Maryland, College Park, MD, 20752

(Dated: January 4, 2011)

Abstract

Boolean networks are discrete dynamical systems in which the state (zero or one) of each network node at time t is updated to a state determined by the states at time $t - 1$ of those nodes that have links to it. Boolean networks with ‘canalizing’ update rules have been of great interest in the modeling of genetic control. A canalizing update rule is one for which the node state at time t is determined by the state at time $t - 1$ of a particular one of its inputs when that input is in its canalizing state. In this paper, we introduce a generalized concept of canalization that we believe offers a significant enhancement in biological relevance, and we obtain a simple general network stability criterion for Boolean networks with generalized canalization for a broad class of network topologies.

Keywords: complex networks, genetic networks, Boolean networks

* pomeranc@umd.edu

Boolean networks have been extensively studied as a model for genetic control of cells [1, 2]. In this framework, the genetic regulatory network is modeled as a directed graph, where links correspond to the influence of one gene on the expression of another. Individual genes are either off or on, represented as 0 or 1, respectively, and the state of a gene at time $t + 1$ is given by a Boolean update function of the states of its inputs at time t . In early analyses, both the network topology and the update functions were assumed to be random. In particular, Kauffman’s $N - K$ network model [1, 3] has received significant study. According to this model, there are N nodes (genes) in the network, each having the same number of input links, K , and the nodes from which these input links originate are chosen randomly with uniform probability. Additionally, the update function determining the time evolution at each node is defined by a random, time-independent, 2^K -entry truth table characterized by the ‘bias’ p , which, as discussed subsequently, is the probability that a one appears as the output of the update function. Using the Hamming distance between two states of the system (i.e., the number of nodes for which the states disagree) as the distance measure, these systems, when large, exhibit both ‘chaotic’ (or unstable) behavior, where the distance between typical initially close states on average grows in time, as well as stable behavior, where the distance decreases. Separating the two is a ‘critical’ regime.

In Ref. [4], we developed an approximate technique for determining the stability of large Boolean networks. A key feature of this work was that it allowed one to investigate the effect of given specific network topology. Numerical experiments were performed [4] exploring such effects as correlation between the number of inputs and outputs at each node [11], assortativity [12], community structure [13], etc., and these experiments yielded results accurately predicted by the theory.

So-called ‘canalizing functions’ are a significant modification of the random truth table model of previous work [14]. Canalizing functions, believed to be biologically relevant [5, 6], are those functions where an argument of the function (the ‘canalizing input’), having a certain value (the ‘canalizing value’), determines the value of the function independent of the values of the other arguments (inputs) [2]. If the canalizing input does not have the canalizing value, the function is determined by the other inputs. (Further refinements can include a hierarchy of canalization [7].) Canalizing functions often stabilize networks that would be unstable in their absence [7, 8]. Shmulevich and Kauffman defined the ‘activity’ [9] of a Boolean variable on a Boolean function, which can be used to quantify the increased

importance of canalizing inputs.

In this paper, we present a generalized model of canalizing behavior that we believe offers an enhancement in the biological relevance of Boolean network models. We use the Shmulevich-Kauffman activity to extend the results of Ref. [4] to the case of networks with canalizing functions. We derive an hypothesized condition under which such networks are stable, and we numerically test this criterion. A significant point is that our stability criterion applies to networks of very general topology.

Boolean networks comprise a state vector $\Sigma(t) = [\sigma_1(t)\sigma_2(t)\dots\sigma_N(t)]^T$, where each $\sigma_i \in \{0, 1\}$, and a set of update functions f_i , such that

$$\sigma_i(t) = f_i(\sigma_{j(i,1)}(t-1), \sigma_{j(i,2)}(t-1), \dots), \quad (1)$$

where $j(i, 1), j(i, 2), \dots, j(i, K_i^{in})$ denote the indices of the K_i^{in} nodes that input to node i , and we denote this set of nodes by $\mathcal{J}_i = \{j(i, k) | k = 1, 2, \dots, K_i^{in}\}$. (In the following discussion, k , which is between 1 and K_i^{in} , is used to label an input to node i , or, equivalently an argument of f_i ; j , which is between 1 and N , refers to the network index of the node corresponding to input k ; $k(i, j)$ and $j(i, k)$ are used to switch between them. Similarly, σ_j is the state of node j , and s_k is the k -th input to f_i .) The number of input links K_i^{in} to node i is called its in-degree, and the number of output links K_i^{out} from node i is called its out-degree. The update function f_i at each node i is usually defined by a truth table, where the table for node i has $2^{K_i^{in}}$ rows, one for each possible set of the states of the K_i^{in} nodes that input to node i , and each input state row is followed by its resulting update output state for node i , thus forming a $2^{K_i^{in}}$ entry output column. The stability of a large Boolean network is defined by considering the trajectories resulting from two close initial states, $\Sigma(t)$ and $\tilde{\Sigma}(t)$. To quantify their divergence, the Hamming distance of coding theory is used: $h(t) = \sum_{i=1}^N |\sigma_i(t) - \tilde{\sigma}_i(t)|$. If the network is stable, on average $h(t) \rightarrow 0$ as $t \rightarrow \infty$. In unstable networks, $h(t)$ quickly increases to $O(N)$, while a ‘critical’ network is at the border separating stability and chaos.

In order to study the stability of $N - K$ Boolean networks, Derrida and Pomeau [14] considered an annealing procedure and calculated the probability that, after t steps, a node state is the same on two trajectories that originated from initially close conditions. (Later authors generalized the Derrida-Pomeau analysis to include variable in-degree [15–17] and joint in-degree/out-degree distributions [20].) In their annealed situation, at *each* time step t the truth table outputs and the network of connections are randomly chosen. The actual

situation of interest, however, is the case of ‘frozen-in’ networks, where the truth table and network of connections are fixed in time. It was hypothesized and later numerically confirmed that, for large networks, results obtained in the analytically tractable annealed situation are the same as those in the analytically intractable frozen situation. In deriving the results of Ref. [4], we used a ‘semiannealed’ procedure in which the network connections were frozen, but truth table outputs were randomly chosen on each time step. Again, the aforementioned hypothesis is very well supported numerically.

The semiannealing procedure used in Ref. [4] independently and randomly reassigned the output elements of the truth table governing node i to be one or zero with probability p_i or $1 - p_i$, respectively. However, canalizing functions do not have this property: if the canalizing input takes its canalizing value in a given row of the table, the probability of a one appearing in the output row is zero (or one). We call this behavior ‘strictly canalizing.’ Considering all possible inputs to have equal probability, we now introduce a generalization of strictly canalizing behavior to the case of ‘quasicanalizing’ inputs, which we define as the case where the probability that a one appears in the output of node i ’s truth table if input k takes value s averaged over all other inputs, $p_i^{(k,s)}$, depends on s . (The average of $p_i^{(k,s)}$ over both values of s , $p_i^* = (p_i^{(k,0)} + p_i^{(k,1)})/2$ is the ‘effective bias.’ This quantity will be discussed in detail below.) Strict canalization with respect to input k , therefore, is the case when $p_i^{(k,s)} = 0$ or 1 when s is the canalizing value. If $p_i^{(k,0)} = p_i^{(k,1)}$, k is a non-canalizing input to i . We call truth tables where all inputs are non-canalizing ‘unstructured,’ and those with any canalizing inputs, strict or quasi-, ‘structured.’ Given our generalized definition of canalizing behavior, in the rest of this paper we formulate a modified semiannealing procedure, assume that this approximates the frozen-in case to derive the stability criterion, and numerically confirm that the stability criterion holds in the frozen-in case with structured truth tables.

A crucial quantity in the theory of Ref. [4] is the ‘sensitivity’ of a node, which is the probability that any change to a node’s inputs causes a change in the node’s output. This quantity treats all nodes equally importantly, however, and this is clearly not appropriate in the case of canalizing functions. A remedy for this is to use the activity [9], of input k on f_i , r_{ik} , which is the probability that the output of f_i changes if *only* input k changes. Presupposing a mapping between the set of $p_i^{(k,s)}$ that describe f_i and the activities r_{ik} and a suitable semiannealing procedure (both of which we derive below), we can now extend the procedure of Ref. [4] to account for canalizing behavior in analyzing Boolean network

stability.

We define the N -dimensional vector $\vec{y}(t)$, where each element $y_i(t)$ tracks the probability that node i differs between two initially close states after t time-steps: $y_i(t) = \Pr[\sigma_i(t) \neq \tilde{\sigma}_i(t)]$. Our goal is to derive an update equation for $y_i(t)$ and perform linear stability analysis on the solution $y_i(t) = 0$. The update equation will be derived under the assumption that the inputs $y_j(t)$ are statistically independent of one another. This assumption holds in the case of locally tree-like topology [4, 10].

Since we are performing linear stability analysis, we can make several simplifying approximations. The probability of d inputs to node i being different between the trajectories of two initially close states is of order $\mathcal{O}(y^d(1 - y^{K_i^{in}-d})) \approx \mathcal{O}(y^d)$. Since in linear stability y is taken to be small, the probability that only input node j to node i is flipped is approximately $y_j(t)$, and the probability that this occurs and leads to a flip in the output of node i is $r_{ik}y_{j(i,k)}(t)$. Thus we get the following approximate evolution equation for small perturbations from the solution $\vec{y}(t) = 0$:

$$y_i(t+1) \approx \sum_{k=1}^{K_i^{in}} r_{ik}y_{j(i,k)}(t) + \mathcal{O}(y^2). \quad (2)$$

This can be written in matrix form after discarding the higher-order terms as

$$\vec{y}(t) \approx R\vec{y}(t-1), \quad (3)$$

where R is the ‘activity matrix’ with elements $R_{ij} = r_{ik}$ if there is a link from j to i ($k = k(i, j)$), and zero otherwise. From this equation, we see that stability is determined by the largest eigenvalue λ_R of this matrix:

$$\begin{aligned} \lambda_R > 1, y = 0 &\text{ is unstable;} \\ \lambda_R = 1, y = 0 &\text{ is critical;} \\ \lambda_R < 1, y = 0 &\text{ is stable.} \end{aligned} \quad (4)$$

Before completing the details of the theory (i.e., specifying how to obtain r_{ij} from $\{p_i^{(k,s)}\}$ and the network topology), we present numerical results testing our derived criterion for the stability of Boolean networks with canalizing truth tables in Fig. 1. We consider two cases of canalization in the truth tables: (a) a varying proportion of nodes have a single, strictly canalizing input (open markers); and (b) each node has a single quasicanalizing

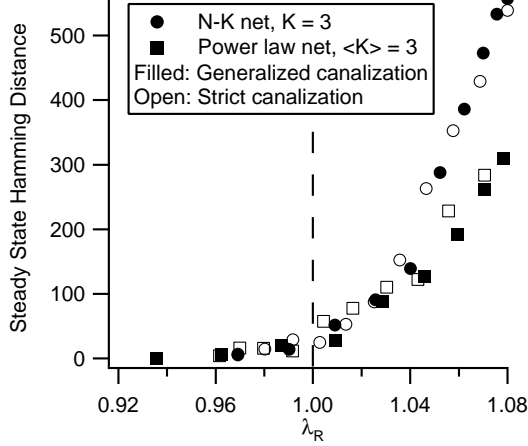


FIG. 1. Steady state Hamming distance vs. λ_R for two different $N = 10^4$ node networks: an $N - K$ network where each node has $K = 3$ inputs (circles), and a network with truncated power law degree distribution with $\langle K \rangle = 3.08$ (squares). λ_R is varied in two ways: either by giving each node a single quasicanalizing input of increasing strength (filled markers), or by giving an increasing proportion of nodes a single, strictly canalizing input (open markers). The predicted transition is at $\lambda_R = 1$ (dashed line).

input of increasing strength (filled markers). We also test the criterion on two network topologies: one on a network with exactly $K = 3$ inputs and outputs per node (circles), and a network where the in- and out-degrees are drawn independently from a truncated power-law distribution: $P(K) \propto K^{-2.8}$ if $K^{min} \leq K \leq K^{max}$, and 0 otherwise. This yields a network with equal values of the in- and out-degree $\langle K \rangle \approx 3$. Both networks have $N = 10^4$ nodes, and the largest eigenvalues of the adjacency matrices of both networks are approximately $\lambda \approx 3$. All nodes in both networks have a uniform effective bias $p_i^* = 0.235$ for every data point. In the absence of canalization, Ref. [4] predicts that the networks would be slightly in the chaotic regime. Each datapoint in Fig. is the average steady-state Hamming distance measurement of 100 different frozen realizations of the truth tables. The steady-state Hamming distance is calculated as the average Hamming distance from time $t = 90$ to $t = 100$ between trajectories that have an initial Hamming distance of 10 (0.1 % of the nodes are flipped).

In the first method of varying λ_R , we increase the proportion of nodes that have a single canalizing input from zero canalizing inputs to each node having one canalizing input. For each node in the network, we choose whether the node will have a canalizing input with

probability p_{can} . When choosing a canalizing input c to node i , we wish to maximize the impact on λ_R , so we choose c that has the minimum value of $K_c^{in} K_c^{out}$. When $p_{can} = 0$, λ_R takes its maximum value; when $p_{can} = 1$, λ_R takes its minimum value.

In the second method of varying λ_R , where each node has a single quasicanalizing input, we choose the canalizing input c as above. When assigning a generalized canalized truth table, we randomly choose the canalizing value v to be zero or one with uniform probability. To vary λ_R , we vary $p_i^{(c,v)}$ from zero to p_i^* : when $p_i^{(c,v)} = 0$, all nodes have a single strictly canalizing input (i.e., it is identical to the case where $p_{can} = 1$ above); when $p_i^{(c,v)} = p_i^*$, c is no longer a canalizing input and the network is identical to the case where $p_{can} = 0$ above. A significant result from Fig. 1 is that while the two networks trace different curves, the two methods of varying λ_R seem to noisily lie on the same curve. We see that the result from our stability criterion (the dashed line) appears to give an extremely good prediction of the transition from the zero Hamming distance state (stability).

Having discussed the condition under which a semiannealed, canalizing Boolean network is stable and given a numerical test for it, we now return to the task of deriving the annealing procedure for the truth tables and, using those results, derive an expression for r_{ij} in terms of the $p_i^{(k,s)}$. We define the appropriate annealing procedure used on the truth tables by specifying the probability that a given output value of f_i is one. A useful quantity in the following analysis is the ‘effective bias’ p_i^* , which is the probability that any truth table output is one, similar to the unstructured case. Letting $L \equiv 2^{K_i^{in}}$ be the number of rows in the truth table, the expected number of ones in the output column of the truth table of node i is $p_i^* L$. For any given arbitrary input k to node i , $L/2$ rows in the truth table have $s_k = 0$ and $L/2$ have $s_k = 1$. By definition of $p_i^{(k,s)}$, the expected number of ones in the output of the truth table with entries that have $s_k = s$ is $p_i^{(k,s)} L/2$. The total expected number of ones is the sum of the expected number of ones when $s_k = 0$ and when $s_k = 1$, which leads to

$$p_i^* = \frac{p_i^{(k,0)} + p_i^{(k,1)}}{2}, \quad (5)$$

Note that, since the expected number of ones does not depend on our choice of k above, $p_i^{(k,0)} + p_i^{(k,1)}$ must be independent of k . This provides a constraint on the set of possible $p_i^{(k,s)}$ values that describe a realizable truth table; the full set of constraints will not be needed for what follows and will be discussed in a followup paper. Non-canalizing inputs have both $p_i^{(k,0)}$ and $p_i^{(k,1)}$ equal to the effective bias by definition, and unstructured truth

tables have $p_i = p_i^*$.

We now derive the probability that a given set of input values to node i , $\{s_1, s_2, \dots, s_{K_i^{in}}\}$, yields an output of one, and we denote this probability $\phi_i(s_1, \dots, s_{K_i^{in}}) \equiv \Pr[f_i = 1 | I_1, \dots, I_{K_i^{in}}]$. Using Bayes' Theorem, we have

$$\phi_i(s_1, \dots, s_{K_i^{in}}) = \frac{\Pr[I_1, \dots, I_{K_i^{in}} | f_i = 1] \Pr[f_i = 1]}{\Pr[I_1, \dots, I_{K_i^{in}}]}, \quad (6)$$

where I_k is the event that the k -th input takes the value s_k (i.e., I_k is the event that $\sigma_j = s_k$, where s_k denotes a specific value, 0 or 1, of the node $j(i, k)$'s state variable σ_j). By definition, $\Pr[f_i = 1] = p_i^*$. Since we are considering an ensemble where every possible input string to f_i has equal probability, $\Pr[I_1, \dots, I_{K_i^{in}}] = 2^{-K_i^{in}}$. We note that since each of the events I_k are independent, $\Pr[I_1, \dots, I_{K_i^{in}} | f_i = 1] = \prod_k \Pr[I_k | f_i = 1]$ and we calculate $\Pr[I_k | f_i = 1]$ again using Bayes' Theorem:

$$\Pr[I_k | f_i = 1] = \frac{\Pr[f_i = 1 | I_k] \Pr[I_k]}{\Pr[f_i = 1]} = \frac{p_i^{(k,s)}(1/2)}{p_i^*}. \quad (7)$$

Using these results in Eq. (6), we obtain

$$\phi_i(s_1, \dots, s_{K_i^{in}}) = (p_i^*)^{1-K_i^{in}} \prod_{k=1}^{K_i^{in}} p_i^{(k,s_k)}. \quad (8)$$

Thus, our new semiannealing procedure, generalized to include canalization, randomly reassigns each output element of the truth table of all nodes at each time according to the probability given by Eq. (8).

Using Eq. (8), we can calculate the activity of input k on f_i , r_{ik} , using the definition that it is the probability that the output of f_i changes if *only* input k changes. We define $\phi_i^{(k,s)} \equiv \phi_i(s_1, \dots, s_{k-1}, s, s_{k+1}, \dots)$ to be a $K_i^{in} - 1$ input function that denotes the probability that the truth table output is one if input k is s given some $K_i^{in} - 1$ element set of other inputs. With this, we calculate the activity as

$$r_{ik} = \langle \phi_i^{(k,0)}(1 - \phi_i^{(k,1)}) + \phi_i^{(k,1)}(1 - \phi_i^{(k,0)}) \rangle, \quad (9)$$

where $\langle \cdot \rangle$ is the average over all $K_i^{in} - 1$ states $s_{k'}$ for $k' \neq k$.

This completes our derivation of our result for assigning an R matrix to a given Boolean network topology and specification of $\{p_i^{(k,s)}\}$. It is this result, along with Eq. (4), that we have used in obtaining the stability threshold plotted as the dashed line in Fig. 1.

In this paper we have presented a probabilistic generalization to canalized functions. Our generalization allows a continuum in the degree of canalization, as opposed to the previous model [5] where an input could only be strictly canalizing or not canalizing at all. We believe that our generalized canalization model could be of enhanced relevance to gene networks. We used this generalized canalization model to define a semiannealing procedure where the update functions of every node in the network are randomly reassigned at each time step, but the network of connections (i.e., the network topology) remained frozen. We employed this semiannealing situation, along with the supposition that it yields results for the frozen case, to derive the condition under which Boolean networks that have canalizing functions are stable in Eq. (4), and we numerically confirmed our supposition. Given the likely prominence of canalizing behavior in gene networks, these results may offer significant input into the understanding of these systems. Furthermore, since our results allow analysis of any specified network (e.g., an experimentally determined network), our stability criterion may eventually, with advances in gene network measurement techniques, allow one to assess the criticality of real genetic networks.

We thank Wolfgang Losert for his comments. This work was partially supported by ONR Grant N00014-07-1-0734.

-
- [1] S.A. Kauffman, *J. Theor. Biol.* **22**, 437 (1969).
 - [2] S.A. Kauffman, *The Origins of Order* (Oxford University Press, New York, 1993).
 - [3] M. Aldana, S. Coppersmith, and L. P. Kadanoff in *Perspectives and Problems in Nonlinear Science*, edited by Kaplan E., Marsden J.E., Sreenivasan K.R. (Springer, Berlin, 2002).
 - [4] A. Pomerance, E. Ott, M. Girvan, and W. Losert, *Proc. Natl. Acad. Sci.* **106**, 8209 (2009).
 - [5] S. A. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, *Proc. Natl. Acad. Sci.* **100**, 14796 (2003).
 - [6] S. E. Harris, B. K. Sawhill, A. Wuensche, and S. A. Kauffman, *Complexity* **7**, 23 (2002).
 - [7] S. A. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, *Proc. Natl. Acad. Sci.* **101**, 17102 (2004).
 - [8] P. Ramo, J. Kesseli, and O. Yli-Harja, *Chaos* **15**, 034101 (2005).
 - [9] I. Shmulevich and S. A. Kauffman, *Phys. Rev. Lett.* **93**, 048701 (2004).

- [10] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. Lett. **100**, 058701 (2008).
- [11] J. G. Restrepo, E. Ott and B. R. Hunt, Phys. Rev. E **76**, 056119 (2007).
- [12] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
- [13] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
- [14] B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986).
- [15] R. V. Solé and B. Luque, Phys. Lett. A **196**, 331 (1995).
- [16] B. Luque and R. V. Solé, Phys. Rev. E **55**, 257 (1997).
- [17] J. J. Fox and C. C. Hill, Chaos **11**, 809 (2001).
- [18] M. Aldana and P. Cluzel, Proc. Natl. Acad. Sci. USA **100**, 8710 (2003).
- [19] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).
- [20] D. S. Lee and H. Rieger, J. Phys. A **41**, 415001 (2008).